

ORIGINAL ARTICLE

Multi-state outcome analysis of treatments (MOAT): application of a new approach to evaluate outcomes in longitudinal studies of bipolar disorder

CL Bowden¹, J Mintz¹ and M Tohen²

Survival analyzes are usually based on a single point in time predefined event. Dissatisfied with this approach to evaluating maintenance treatment outcomes, we developed the Multi-state Outcome Analysis of Treatments (MOAT) methodology using a combined database from two FDA registration studies of lamotrigine, lithium and placebo. MOAT partitions total survival time into clinically distinct periods operationally defined by cutpoints on rating scales. For bipolar disorder (BD), the clinical states are remission, subsyndromal and syndromal mania, mixed states or depression. MOAT results can be crossed with information about tolerability and functioning to yield an outcome system integrating efficacy and tolerability. As found in the original analysis, both drugs were associated with longer time in study compared with the placebo. MOAT supplements this by finding that both drugs increased the time remitted compared with placebo. However, a substantial amount of time in all three treatments was spent in subsyndromal depression. Time with manic symptoms was reduced with lithium, but not lamotrigine. Patients on placebo neither benefitted nor had adverse effects from the assignment but experienced more syndromal levels of symptoms and were terminated from the study sooner than either drug treated group. Lithium was associated with both benefit in time manic and worse tolerability compared with placebo. In summary, lamotrigine was associated with limited therapeutic benefit but not harm; lithium with both benefit and harm; and placebo with neither. MOAT describes not only quantity but also quality of time spent in longitudinal studies, providing a more clinically informative picture than Kaplan–Meier survival analysis.

Molecular Psychiatry (2016) **21**, 237–242; doi:10.1038/mp.2015.21; published online 17 March 2015

INTRODUCTION

A limitation of traditional survival analysis of time to some clinically interesting event is the inattention to what happens between time zero and the specified outcome, including subsyndromal symptoms and problems with tolerability. Kaplan–Meier (KM) survival techniques provide only a single efficacy result of time to event (relapse, intervention, discontinuation). This project was supported by an NIMH-funded grant to develop tools that incorporate data on both efficacy and safety applied to time spent in primary clinical states of bipolar disorders (BD) to guide investigation of illness trajectories and treatment selection.

Our initial effort was to adapt the Quality-adjusted Time Without Symptoms or Toxicity method to BD to overcome this limitation.^{1,2} Developed and applied in cancer chemotherapy trials, Quality-adjusted Time Without Symptoms or Toxicity method divides survival time into periods defined by the presence or absence of symptoms or toxicity. A central feature of Quality-adjusted Time Without Symptoms or Toxicity method is the use of weights that yield a single estimate of ‘quality-adjusted’ survival time. These estimates count a day as less than a day when quality-of-life is diminished by toxicity or recurrence of illness. Cox *et al.*³ noted, ‘This approach, although attractive, has clear limitations. It is preferable to avoid the explicit combination of quality and length of life and instead to present them as multiple end points of a trial. These end points will have to be formally or informally

combined for making treatment decisions, but responsibility for this is left to the clinician and the patient.’

We agree with Cox *et al.*³ that a different approach was needed for BD and other waxing and waning chronic diseases. In BD, subsyndromal fluctuations of mood that impair functionality and quality-of-life are more the rule than exception. For example, a bipolar patient who develops sufficient symptoms to warrant intervention usually has resolution of the state and proceeds to better symptomatic and functional states over time.^{4,5} We did not concur that the task of integrating quality and length of survival should be left to subjective judgments of clinician and patient. Multi-state Outcome Analysis of Treatments (MOAT) is our response to that challenge. We developed MOAT to more fully capture the actual course of maintenance treatments in BD. We report the principles of its development and examples of its performance in re-analyses of data from two registration studies of lamotrigine in comparisons with lithium and placebo in maintenance treatment of BD.^{6,7}

MATERIALS AND METHODS

Source of data: the GSK study

For the analyzes presented here, we accessed data from two studies previously published separately^{6,7} and with the samples pooled.⁸ Details of the samples can be found in the original publications. One of the studies recruited the recently depressed⁶ and the other recently manic⁷ patients.

¹Department of Psychiatry, Division of Mood and Anxiety Disorders, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA and ²Department of Psychiatry, University of New Mexico Health Sciences Center, Albuquerque, NM, USA. Correspondence: Dr CL Bowden, Department of Psychiatry, 7703 Floyd Curl Dr, Division of Mood and Anxiety Disorders, University of Texas Health Science Center at San Antonio, San Antonio, TX 78229-3900, USA.
E-mail: bowdenc@uthsca.edu

Received 21 April 2014; revised 2 December 2014; accepted 19 December 2014; published online 17 March 2015

The two studies were similar in all respects, including clinical criteria for randomization, assessments, outcome criteria, duration, criteria for censoring and overall analytical plan. After an 8- to 16-week open-label phase, during which lamotrigine was initiated as adjunctive therapy or monotherapy and other psychotropic drugs were discontinued, patients were randomized to receive lamotrigine, lithium or placebo in a 76-week double-blind maintenance phase. Prior to randomization, patients were required to have received lamotrigine alone for a minimum of 1 week and to have maintained a CGI Severity scale specific to BD score of 3 or less for at least 4 continuous weeks. Lamotrigine was dosed at 100–400 mg per day, and lithium was titrated to serum levels of 0.8–1.1 mEq l⁻¹. We were able to replicate the published primary outcome results of both studies, confirming that the data that we analyzed by MOAT represented the same subjects and measures. This report is based on $n=578$ patients, 224 assigned to lamotrigine, 165 to lithium, and 189 to placebo.

Operational definition of clinical states and pivotal decision points

The unit of analysis in MOAT is the *period*. MOAT partitions the total time a patient is observed during a study into one or more discrete periods, each representing the duration of time spent in one of several operationally defined clinical states. In principle, these states could be defined by measures from any domain of interest. We chose well-known measures of manic and depressive symptoms as the primary basis for defining periods, coupled with information about adverse events (AEs).

We adapted published clinical state criteria as recommended by the International Society for Bipolar Disorders to define the mood states.⁹ For this developmental study of MOAT we used the two symptom scales applied in the lamotrigine studies: the 17-item version of the Hamilton Depression Rating Scale (HDRS)¹⁰ and the 11-item Mania Rating Scale (MRS),¹¹ to define subsyndromal or syndromal symptomatology, with further separation by predominant subtype (depressive, manic, mixed). This seven-state classification system is presented in Table 1. Note that syndromal depressive states can include subsyndromal levels of manic symptoms. Conversely, syndromal mania can also include subsyndromal depressive symptoms. Subsyndromal mixed states are defined by intermediate levels on both scales, and syndromal mixed states require high scores on both instruments.

Defining periods and their duration

We sought a consistent approach to define the time points wherein a clinical state begins and ends. Table 2 presents hypothetical data for one patient in the data format often used in clinical trials. The table displays four hypothetical data records, one for each of the four assessments, with scores on the two assessment instruments. These data do not represent the actual assessment schedule in our data set. The assessment days were chosen for ease of explication of the methodology. In this case, we suppose a patient was assessed on days 1, 5, 9 and 15 with the HDRS and MRS. On the basis of these scale data, each record is classified into one of the clinical states using the system described in Table 1. This patient was classified as remitted on Day 1, in subsyndromal depression on days 5 and 9, and remitted again when assessed on day 15.

Table 3 illustrates how the 15 days summarized by these four hypothetical records would be used to define three *periods* for MOAT analysis. Lacking precise information as to when transitions occur, we assume that changes in clinical state occur at the midpoint between assessments. For the data in Table 2, the transition from remitted to subsyndromal depression is assumed to have occurred midway between days 1 and 5 at day 3, and the transition back to remitted between days 9 and 15 on day 12. With evaluations scheduled from 7 to 28 day intervals all midway dates ranged from 3–14 days. This procedure eliminated possible bias consequent to patient or investigator guessing.

Most periods have a known end point because the patient transitions from one state to another. In that case, calculation of durations of the periods is straightforward (end day minus start day plus one, so a period beginning on day 4 and ending on day 7 is: 7–4+1=4 days). The same is true if patients are observed throughout the entire study period so that the last day is known. However, if the patient discontinues or is lost to follow-up, the true duration of the final state is not known. In the language of survival analysis, that observation is censored. Given the common assumption that censoring is uninformative, the convention is to impute the mean of all known longer event times for a censored observation. MOAT does that as well, but estimates of state durations are state-specific in MOAT. If a patient drops out in a state of syndromal depression, for

Table 1. Clinical state classification system (seven categories)

MRS	Hamilton Depression Rating Scale: 17 items		
	0–6	7–15	16+
0–5	Remitted	SS depression	SYN depression
6–15	SS mania	SS mixed	SYN depression
16+	SYN mania	SYN mania	SYN mixed

Abbreviations: MRS, Mania Rating Scale; SS, subsyndromal; SYN, syndromal.

Table 2. Hypothetical symptom data for one patient

Assessment day	State	HDRS	MRS
1	Remitted	2	1
5	Subsyndromal depression	8	2
9	Subsyndromal depression	12	2
15	Remitted	1	0

Abbreviations: HDRS, Hamilton Depression Rating Scale; MRS, Mania Rating Scale; SS, subsyndromal; SYN, syndromal.

Table 3. Symptom records from Table 2 recoded as Multi-state Outcome Analysis of Treatments periods

Period	State	Start day	End day	Duration
1	Remitted	1	3	3
2	Subsyndromal depression	4	12	9
3	Remitted	13	15	3
			Total	15

example, the estimated duration of that censored period will be based only on other periods of syndromal depression with longer durations. MOAT treats the longest observed time as an event,^{12,13} and constrains imputation of censored observations so that the total time contributed by any individual patient does not exceed any limit set on the study duration. The latest actual end point of any period was 357 days, thus imputed end points for any censored MOAT period that extended beyond this were truncated at 357 days. This step is termed restricted mean event times in survival analysis.^{13,14}

Estimating mean durations and their standard errors

Calculation of the mean duration of each state is also straightforward. Adding up the durations across periods produces the total number of days in each state for each patient. The minimum duration is zero for any patient who spend no time in a state. Some states occur infrequently, resulting in positively skewed distributions. Our MOAT programming uses bootstrap estimates of the standard errors of the mean durations for significance testing, taking the patient as the resampling unit. The bootstrap estimates of standard errors minimizes the role of assumptions in statistical testing. Overall tests of significance were done with Cochran's *F*-test, a modification of his *Q* statistic for testing homogeneity of multiple parameters using the bootstrap estimates of the standard errors and assuming unequal precision.¹⁵ Our experience has been that significance testing using the generalized linear model (e.g., SAS GENMOD) produces almost identical results. For purposes of this methods development research, we report unadjusted *P*-values ≤ 0.05 as significant. The programming was done using the SAS statistical system (version 9.3, SAS Institute, Cary, NC, USA). Procedural details and download of SAS 9.3 macro code to conduct MOAT analyzes are available at: <https://delta.uthscsa.edu/moat>.

In contrast with the traditional survival analysis, MOAT is both compatible with and encouraging of retention of subjects, as well as clinically safe. Study designs that retain high proportions of subjects allow insights into illness course and drug effects that are lost in standard

survival analyzes that follow participants only to the time of first event. Even the increasingly popular mixed effects statistical methods require restrictive and ultimately untestable assumptions about the randomness of the unobserved data. The two source studies enrolled all patients who in the open-label phase maintained a CGI severity score ≤ 3 for at least 4 weeks of lamotrigine monotherapy. At each time point in the MOAT analyzed study, each patient occupies one, and only one, of the seven clinical states. MOAT thus provides precisely operationalized definitions of each subject's clinical states over the course of the randomized trial.

AEs and tolerability

We wanted AEs to represent side effects, not symptoms of the illness. Therefore, events that were clearly related to the primary symptom outcomes (e.g., depression, mania, hospitalization) were excluded. However, AEs that were diagnostically ambiguous were not excluded, for example, 'lack of energy,' which might or might not indicate depression. AEs were coded using a four-level ordinal system based on the most serious event for each participant (called MaxAE below). The most severe category was assigned to an AE resulting in permanent discontinuation of the study drug. The next category was a temporary discontinuation of study medication, dosage reduction or other action. The third category was for participants with AEs noted, but no action taken. The lowest level was assigned to patients with no recorded AEs. A second measure was derived by counting the number of AEs recorded during the study (called #AE).

Integrating symptom states and drug tolerability with latent class analysis

To obtain an outcome classification system that integrated symptom states as defined by MOAT with measures of drug tolerability, we utilized latent class analysis (LCA) as implemented in SAS PROC LCA, Version 1.2.7.¹⁶ PROC LCA is a type of cluster analysis that groups the patients into subgroups whose members have similar response profiles. The classification variables were MOAT estimates of symptom state durations and the measures based on AEs. PROC LCA requires that the classification variables be categorical, so all of the variables were dichotomized at their median values. PROC LCA reports likelihood-based information criteria to guide the decision about number of groups to retain, but as with factor analysis, both clinical judgment and statistical considerations are used.

RESULTS

MOAT multi-state analyzes of symptoms

Table 4 summarizes the MOAT multi-state analyzes of duration of the seven symptom states and some summary totals. Bold type indicates a significant omnibus *F*-test, which when significant was followed-up with pairwise tests reported in the Note based on the bootstrap estimates of the standard errors. As noted in the Table, the reported *P*-values for the omnibus tests are unadjusted for multiple testing, which we believe is appropriate for exploratory analyzes such as these. Conservative Bonferroni-adjusted values can be obtained by multiplying the reported *P*-values in the upper section by seven tests performed, and those in the lower section, which are based on sums of those above, by four.

Consistent with the original published findings, total time in study was significantly longer for both the active drugs than placebo.⁸ This was primarily owing to longer time remitted on both drugs, with days remitted making up ~59% of the total study time on both active drugs and ~52% on placebo. Lithium was associated with fewer days with subsyndromal or any manic symptoms than placebo. Across all three treatments, the time spent in subsyndromal depression is notable, representing 23% of placebo study days and 24% for each active drug. Neither lithium nor lamotrigine differed from placebo on days with subsyndromal or syndromal depression or mixed states.

Integrated outcome profiles identified with LCA

LCA identified six outcome subgroups (Table 5). The analysis was based on the measures of symptoms and AEs. The symptom

Table 4. MOAT estimates of mean state durations \pm Bootstrap standard error.

State	Mean duration \pm standard error			Cochran's F-value (P-value)
	PBO	Lithium	LTG	
Remitted	99.0 \pm 8.2	131.3 \pm 9.7	137.0 \pm 7.4	6.38 (0.002) ^a
SS mania	11.8 \pm 2.5	5.4 \pm 1.1	8.5 \pm 1.3	3.70 (0.026) ^b
SS depression	44.1 \pm 4.9	53.4 \pm 6.6	55.7 \pm 4.9	1.52 (0.220)
SS mixed	7.2 \pm 1.9	6.0 \pm 2.0	6.6 \pm 1.5	0.09 (0.913)
SYN mania	6.5 \pm 1.3	3.8 \pm 1.2	5.0 \pm 1.0	1.13 (0.325)
SYN depression	16.4 \pm 2.3	17.3 \pm 2.9	16.0 \pm 1.9	0.08 (0.924)
SYN mixed	5.9 \pm 1.5	4.2 \pm 1.5	4.4 \pm 1.2	0.40 (0.667)
Totals				
SYN/SS mania	18.3 \pm 2.8	9.2 \pm 1.8	13.5 \pm 1.8	3.91 (0.021) ^c
SYN/SS depression	60.5 \pm 5.4	70.7 \pm 8.1	71.6 \pm 5.5	1.18 (0.308)
SYN/SS mixed	13.1 \pm 2.8	10.2 \pm 2.8	11.0 \pm 2.0	0.29 (0.746)
Total days	190.8 \pm 8.9	221.4 \pm 9.5	233.2 \pm 8.9	6.00 (0.003) ^d

Abbreviations: LTG, lamotrigine; MOAT, Multi-State Outcome Analysis of Treatments; PBO, placebo; SS, subsyndromal; SYN, syndromal. Note: Cochran's *F*-value has a numerator *df*=2 and the denominator *df* range from 351 to 377. *P*-values are unadjusted for multiple testing. Pairwise comparisons: ^aLithium-PBO *z*=2.53, *P*=0.011; LTG-PBO *z*=3.43, *P*=0.0006. ^bLithium-PBO *z*=-2.38, *P*=0.018. ^cLithium-PBO *z*=-2.71, *P*=0.007. ^dLithium-PBO *z*=2.35, *P*=0.019; LTG-PBO *z*=3.37, *P*=0.0008.

state variables used were the MOAT duration estimates, transformed to percentages of total time in study. Additional measures were the severity of the worst AE and number of AEs. *SYN Dropout* meant that the patient left the study in a syndromal state; *Dose Stopped* meant the study drug was terminated because of AEs.

Entries in the body of the table are the proportion in each of the latent classes who scored above the median on each of the indicators. For example, 73% of the patients assigned to the first latent class (see column 1, 'Remitted without AE/side effects') were above the median total time in study, all of them (100%) were above the median in time remitted and none were above average in frequency or severity of AEs. Grayed out entries are not significantly related to class membership as determined by the χ^2 tests using a Bonferroni-adjusted significance criterion of *P*=0.05/66=0.0008 based on performing 66 tests (11 measures \times 6 classes).

The first three columns of Table 5 define subgroups of patients, all of whom are low in AE severity and frequency. The three rightmost columns are groups that are above the median in AE severity and frequency. In these last three subgroups, for example, between 17–25% of patients had drug stopped prematurely compared with only 10% in the total sample who stopped drug prematurely. In contrast, none of the patients in the first three subgroups had this outcome. Within the halves of the table (left and right, defined by the absence or presence of AEs), the subgroups are ordered by symptom severity and represent predominantly good (remitted), fair (subsyndromal) and poor (syndromal) symptom outcomes respectively.

The rows at the bottom of the table summarize latent class assignment as a function of medication, and statistical tests of the association of class membership with medication condition. Bold type highlights the classes that are significantly associated with medication assignment. Medication is a significant predictor of group membership in three of the six latent classes. Lamotrigine

Table 5. Percent in each of six latent class groups scoring above median on measures of symptoms and AEs

Measure	Latent classes					
	Classes without AE/side effects			Classes with AE/side effects		
	Remitted	SS	SYN	Remitted	SS	SYN
Total time in study	73%	56%	14%	71%	61%	17%
Time remitted	100%	37%	21%	100%	29%	10
Time SS	0%	100%	38%	5%	99%	47%
Time SYN	12%	34%	100%	9%	52%	100%
Time manic	21%	34%	45%	23%	30%	28%
Time depressed	1%	77%	60%	0%	87%	71%
Time mixed	5%	29%	43%	6%	38%	44%
Severity of worst AE	0%	0%	0%	100%	100%	100%
Number of AEs	0%	0%	0%	100%	100%	100%
SYN dropout	0%	2%	100%	4%	0%	100%
Drug stopped	0%	0%	0%	25%	17%	21%
Class membership— <i>n</i> (%)						
LTG	59 (26%)	39 (17%)	27 (12%)	20 (9%)	47 (21%)	32 (14%)
Lithium	30 (18%)	21 (13%)	20 (12%)	37 (22%)	34 (21%)	23 (14%)
PBO	28 (15%)	37 (20%)	49 (26%)	19 (10%)	31 (16%)	25 (13%)
Significance of medication differences—Wald's χ^2 (<i>P</i>)						
Overall χ^2 <i>df</i> =2	8.9 (0.012) ^a	3.0 (0.22)	16.9 (0.0002) ^b	16.6 (0.0003) ^c	1.6 (0.45)	1.7 (0.42)

Abbreviations: AE, adverse event; LTG, lamotrigine; PBO, placebo; SS, subsyndromal; SYN, syndromal. Pairwise contrasts: ^aLithium-PBO $\chi^2 = 0.73$, *P* = 0.39; LTG-PBO $\chi^2 = 8.0$, *P* = 0.005; Lithium-LTG $\chi^2 = 3.54$, *P* = 0.06. ^bLithium-PBO $\chi^2 = 10.2$, *P* = 0.0014; LTG-PBO $\chi^2 = 12.6$, *P* = 0.0004; Lithium-LTG $\chi^2 = 0.00$, *P* = 0.98. ^cLithium-PBO $\chi^2 = 9.7$, *P* = 0.002; LTG-PBO $\chi^2 = 0.15$, *P* = 0.70; Lithium-LTG $\chi^2 = 13.0$, *P* = 0.0003.

increased the likelihood of being in Group 1 (good symptom outcome without AEs) by about 50%. Placebo roughly doubled the likelihood of being in Group 3 (syndromal symptoms without AEs). Lithium roughly doubled the likelihood of being in Group 4 (good symptom outcome but with high AEs). In summary, lamotrigine was associated with therapeutic benefit but not harm; lithium with benefit and harm; and placebo with neither benefit nor harm (Figure 1).

DISCUSSION

Conventional survival analysis addresses questions about the timing and occurrence of events. To be sure, prevention of events such as relapse or death is important, but the quality of time spent in maintenance treatment is at least as important as whether or not these events occur. In BD, subsyndromal fluctuations of mood that impair functionality and quality-of-life are more the rule than the exception. Survival analysis says nothing about the quality of the time until target events happen, or the experience of the many persons who never have the event. MOAT analyzes applied to combined data from the two registration studies of lamotrigine, lithium and placebo in maintenance treatment of BD revealed important clinical trajectory information that neither survival analysis nor mixed effects regression can see. For example, in all three treatment conditions only 50–60% of the total survival time was remitted, and a considerable amount of time (roughly 25%) was spent with subsyndromal depressive symptoms. A variant survival method, competing risk models, is primarily concerned with sampling biases consequent to dropout. Like survival models in general, they are ‘event-focused’ models. The innovation of MOAT is that it is not an event-focused application of survival analysis. MOAT analyzes are really not about ‘events.’ Rather, MOAT describes the duration of time spent in various clinical states. Statistical power in survival analysis is a direct function of the number of events. MOAT analyzes are strengthened if all patients continue to be assessed regardless of which states or how many they experience.

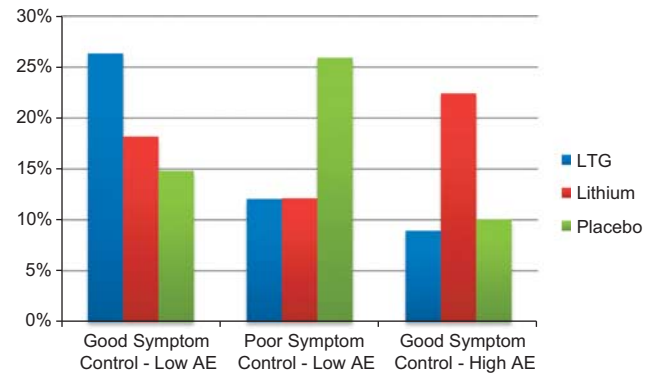


Figure 1. Integrated symptom-adverse events (AE) outcomes. LTG, lamotrigine.

Those statistics provide a realistic perspective on the actual experience of maintenance treatment of BD. Furthermore, MOAT confirmed that both active drugs increased not just total time, but time remitted relative to placebo. Lithium was associated with less time with manic symptoms than either placebo or (non-significantly) lamotrigine. Combining symptom and tolerability data into an integrated outcome profile suggested three different specific medication effects. Lamotrigine increased the likelihood of having a good symptom outcome without side effects by about 50%, although even with lamotrigine that outcome profile only occurred for about 25% of patients. Lithium also increased the likelihood of having a good symptom outcome relative to placebo, but coupled with problems of adverse effects, tolerability, and increased likelihood of having to stop the study medication. Placebo increased the likelihood of having a poor symptom outcome without AEs.

MOAT yields statistically reliable and well powered data on several components of outcome, for example, time in subsyndromal depression, time in depression that is either syndromal or subsyndromal in severity. Investigators utilizing MOAT will have

the responsibility of identifying *a priori* one primary outcome, and several secondary outcomes. Such decisions should be relatively easily made for the majority of studies which are principally intended to strengthen evidence of effectiveness of regimens, both aimed at psychiatrists and persons with the disease of interest. For use in registration studies, the decision would be settled through conference with regulatory agency staff, principally because a successful study determines the label for use of the drug/regimen. By extension, for MOAT-designed studies associated with hypotheses regarding domains of behavior or biological systems, investigators would need to take into account in planning analysis of the clinical state variables any evidence of association with the biological system under study, for example, calcium signaling, family history of the illness or genomic and epigenetic systems measurable in biological samples.

We think these kind of findings may help providers and patients as they consider the maintenance treatments. If patients have realistic expectations, they are more likely to be adherent. Comparative effectiveness trials in BD should also benefit from MOAT analysis.¹⁷ Although the development of MOAT has been in the context of BD, other chronic disorders that require maintenance treatment could benefit from MOAT methodologies. By assessing benefits and harms simultaneously, MOAT increases the granularity of benefits assessment.

Survival analyzes are not necessarily limited to study of a single event, although they typically are. Statistical methods exist for study of multiple events of either the same (multiple relapses) or different (efficacy, tolerability) types. Investigators commonly perform multiple survival analyzes, switching the roles of which is the target event and which are considered censored. Survival analyzes are often supplemented with longitudinal mixed effects analyzes of group means over time at fixed assessment points.¹⁸ These have their value, but longitudinal mixed effects analyzes produce group averages and do not yield estimates of the proportion of time that a typical patient spends in various clinical states. In essence, repeated measures analyzes present a series of snapshots of groups whose membership is constantly changing over time. Neither of these approaches looks at multivariate clinical states or the integration of efficacy and tolerability measures at the level of the individual patient.¹⁹

Some study goals may be incompatible with MOAT as an organizing methodology for conduct and analysis of the research. A study which anticipates low proportions of enrolled subjects completing the trial would be unsuited, as the value of following a patient through the several illness states associated with the disease is intrinsic to the goal of parsing out time in major clinical states. Pharmaceutical company-sponsored studies have principally been developed for the purpose of establishing evidence that a drug is efficacious and that word is typically narrowly defined. So long as the FDA only requires superiority on a binary outcome, rather than time in state data, MOAT is unlikely to be utilized by pharmaceutical corporations. The US FDA has traditionally been slow to change outcome requirements for regulatory approval.

Our decision to pool samples of recently depressed and recently manic patients enrolled in the same research protocol highlights this difference between MOAT and conventional analyzes. Clinical efficacy trials typically impose strict sampling criteria with the goal of obtaining highly homogeneous samples in order to focus on specific drug effects, such as efficacy for mania or prevention of depressive relapse. In contrast, MOAT accommodates study of more heterogeneous samples so that the range of treatment effects, both positive and negative, on a range of symptom outcomes can be elucidated.

MOAT, as well as other long-term methodologies, generally require large samples. High impact randomized intervention studies in BD conducted without a pharmaceutical company funding and administration have been consequent to academic

centers coming together to design and conduct a single study, or a group of studies over time. These groups have often applied novel methodologies for some or all study designs and analyzes. Given the more clinically relevant information that would be provided by MOAT-type studies at costs equivalent to KM analytic designs, some application of MOAT has pragmatic appeal. The cost per subject of studies in these consortia, usually borne by external grants, are substantially lower on a per patient cost than pharmaceutical industry studies. The cost advantages are in-part consequent to lower overhead costs, non-profit financial plans and, in some cases, utilization of insured medical plans for a portion of the enrolled patients' study expenses. Examples in the United States are the NIMH Collaborative Clinical and Psychobiological Programs on Depression, Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD), bipolar trials network, Stanley foundation network and CHOICE. Similar groups based in Europe have been formed and successfully conducted.

Application of MOAT could aid in translational clinical research studies of complex diseases. For example, CACNA1C genetic variants have been identified as the most consistently significant genome-wide risk factors for BD.^{20–22} To date, CACNA1C studies focused on differentiating BD from other groups cross-sectionally. Translational research would benefit from additional tools to study phenomenology, course of illness and treatment effects. Longitudinal studies utilizing MOAT analyzes could help identify biological markers associated not only with having a disorder but with different patterns of illness expression or treatment response.

One limitation of MOAT procedures is that state durations must be estimated if they are truncated by study discontinuation. In particular, the durations of relapse (syndromal) states are underestimated when the study design terminates assessments, as soon as exacerbations occur. Designs that follow patients after relapse until the exacerbation resolves would be particularly well suited to MOAT. KM analysis has, to the contrary, often been associated with termination of a high proportion of enrolled subjects in the first few weeks of a study.²³ Accuracy in determining state durations is obviously enhanced when study assessments are relatively regular and frequent to minimize errors due to recall.

The cutpoints used here to define the clinical states (Table 2) were substantially based on the guidelines published by a task force of the International Society of Bipolar Disorders.⁹ A finer-grained approach could be taken. Validated instruments exist²⁴ to yield estimates of severity in multiple domains in BD, for example, anxiety, irritability, mania, depression and so on.²⁵ Of course, the decision on the number of clinical states should be established cognizant of the study objectives and/or the average severity of the particular patient sample and disease under study.

Integration of data about symptoms and tolerability depends on good measures of both. Symptom status is routinely assessed in maintenance studies, but definitions and recording procedures of tolerability and safety data are highly variable from one study to another. Details needed to ascertain the degree of tolerability are often lacking.²⁶ The HDRS and MRS scales used in the lamotrigine studies inadequately address several fundamental bipolar illness features, for example, oversleeping, affective lability, impulsivity

Table 6. Background characteristics

	Study M	Study D
Age (years)	41.1	43.4
Duration ill (weeks)	18.5	21.3
Number of depression episodes in the last 3 years	1.7	3.8
Number of manic episodes in the last 3 years	3.0	2.2
Number of mixed episodes in the last 3 years	0.6	0.4

and anxiety. Use of a more comprehensive symptom scale would strengthen the validity and generalizability of MOAT analyzes.²⁷

MOAT addresses several misconceptions about what KM-survival analyzes can achieve. One is that KM analysis assesses the ability of a treatment to provide mood stability. KM analysis typically provides no information about either mood stability or subthreshold symptom intensity.¹⁹ Similarly, a patient experiencing early relapse could, after the episode, maintain a stable pattern free of episodes or develop subsyndromal symptoms, neither of which are captured by the KM methodology.

The samples analyzed here come from the only registration program to date for a now approved drug which from early conception of the phase-2 studies anticipated a combined analysis of maintenance phase treatment of patients enrolled based on having experienced a current or recent depression (Study D) or, conversely a current manic/hypomanic episode (Study M). This required planning essentially identical protocols regarding other inclusion-exclusion criteria and study interventions and assessments. In most aspects clinical, illness course and demographic features were quite similar in the two separate cohorts (Table 6). The only substantial change in study execution, once the two separate studies were underway, was to end enrollment of patients into Study M before initial target enrollment was met, a decision made in order to prioritize impact of the study on depressive outcomes while managing the overall costs. Thus more patients in the combined analysis were from study D. No differences were present for gender, age of onset of first depressive or manic episode, total number of mood episodes or CGI Severity score at screening. We noted in the combined study manuscript that... 'the large sample size of this combined database provided significant advantages over the individual studies, including increased statistical power to detect treatment differences.'

We hope to stimulate others to revisit methods for analyzing the longitudinal studies. Used in conjunction with survival analysis, mixed effects regression models and other statistical approaches, MOAT could support analyzes pertinent to effectiveness and personalized treatment. Such analyzes could strengthen the generalizability of maintenance study results for clinical practice and inform treatment guidelines for BD and other disorders.

CONFLICT OF INTEREST

CLB has received research support from Elan and participated in a Data Monitoring Board for Takeda. JM declares no conflict of interest. MT was a full time employee at Lilly (1997–2008). He has received honoraria from, or consulted for, Abbott, AstraZeneca, Bristol Myers Squibb, GlaxoSmithKline, Lilly, Johnson & Johnson, Otsuka, Merck, Sunovion, Forest, Geodon Richter Plc, Roche, Elan, Alkermes, Lundbeck, Teva, Pamlab, Wyeth and Wiley Publishing and his spouse was a full time employee at Lilly (1998–2013).

ACKNOWLEDGMENTS

Data for these analyzes were generously provided by GlaxoSmithKline. This work was supported by a grant from NIMH (RC1MH088431). Analyses of Registration Bipolar Prophylaxis Trials to Develop New Study Designs to TM, with CLB and JM as the co-investigators.

REFERENCES

- 1 Glasziou PP, Simes RJ, Gelber RD. Quality adjusted survival analysis. *Stat Med* 1990; **9**: 1259–1276.
- 2 Gelber RD, Cole BF, Gelber S, Aron G. Comparing treatments using quality-adjusted survival: the Q-Twist method. *Am Stat* 1995; **49**: 161–169.
- 3 Cox DR, Fitzpatrick R, Fletcher AE, Gore SM, Spiegelhalter DJ, Jones DR *et al*. Quality-of-life assessment: can we keep it simple? *JSTOR* 1992; **155**: 40.
- 4 Tohen M, Wateraux CM, Tsuang MT. Outcome in mania. A 4-year prospective follow-up of 75 patients utilizing survival analysis. *Arch Gen Psychiatry* 1990; **47**: 1106–1111.
- 5 Judd LL, Akiskal HS, Schettler PJ, Endicott J, Maser J, Solomon DA *et al*. The long-term natural history of the weekly symptomatic status of bipolar I disorder. *Arch Gen Psychiatry* 2002; **59**: 530–537.
- 6 Calabrese JR, Bowden CL, Sachs G, Yatham LN, Behnke K, Mehtonen OP *et al*. A placebo-controlled 18-month trial of lamotrigine and lithium maintenance treatment in recently depressed patients with bipolar I disorder. *J Clin Psychiatry* 2003; **64**: 1013–1024.
- 7 Bowden CL, Calabrese JR, Sachs G, Yatham LN, Asghar SA, Hompland M *et al*. A placebo-controlled 18-month trial of lamotrigine and lithium maintenance treatment in recently manic or hypomanic patients with bipolar I disorder. *Arch Gen Psychiatry* 2003; **60**: 392–400.
- 8 Goodwin GM, Bowden CL, Calabrese JR, Grunze H, Kasper S, White R *et al*. A pooled analysis of 2 placebo-controlled 18-month trials of lamotrigine and lithium maintenance in bipolar I disorder. *J Clin Psychiatry* 2004; **65**: 432–441.
- 9 Tohen M, Frank E, Bowden CL, Colom F, Ghaemi SN, Yatham LN *et al*. The international society for bipolar disorders (ISBD) task force report on the nomenclature of course and outcome in bipolar disorders. *Bipolar Disord* 2009; **11**: 453–473.
- 10 Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960; **23**: 56–62.
- 11 Endicott J, Spitzer RL. A diagnostic interview: the schedule for affective disorders and schizophrenia. *Arch Gen Psychiatry* 1978; **35**: 837–844.
- 12 Datta S. Estimating the mean life time using right censored data. *Stat Methodol* 2004; **2**: 5.
- 13 Zhong M, Hess KR. Mean survival time from right censored data. *Collect Biostat Res Arch* 2009; **66**: 1–5.
- 14 Efron B. The two sample problem with censored data. *Berkeley Symp Math Stat Prob* 1967; **4**.
- 15 Cochran W. The combination of estimates from different experiments. *Biometrics* 1954; **10**: 101–129.
- 16 Lanza ST, DJ J, Huang L, Xu S, Collins LM. Software Proc LCA & Proc LTA users' guide (Version 1.2.7). 2011, University Park: The Methodology Center: Pennsylvania, USA.
- 17 Nierenberg AA, Sylvia LG, Leon AC, Reilly-Harrington NA, Ketter TA, Calabrese JR *et al*. Lithium treatment -- moderate dose use study (LiTMUS) for bipolar disorder: rationale and design. *Clin Trials* 2009; **6**: 637–648.
- 18 Baethge C, Schlattmann P. A survival analysis for recurrent events in psychiatric research. *Bipolar Disord* 2004; **6**: 7.
- 19 Raudenbush SW, Bryk AS. *Hierarchical Linear Models: Applications and Data Analysis Methods*, (2nd ed.) SAGE Publications: University of Chicago, 2002.
- 20 Roussos P, Giakoumaki SG, Georgakopoulos A, Robakis NK, Bitsios P. The CACNA1C and ANK3 risk alleles impact on affective personality traits and startle reactivity but not on cognition or gating in healthy males. *Bipolar Disord* 2011; **13**: 250–259.
- 21 Ferreira MA, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L *et al*. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet* 2008; **40**: 1056–1058.
- 22 Sklar P, Smoller JW, Fan J, Ferreira MA, Perlis RH, Chambert K *et al*. Whole-genome association study of bipolar disorder. *Mol Psychiatry* 2008; **13**: 558–569.
- 23 Suppes T, Vieta E, Liu S, Brecher M, Paulsson B. Maintenance treatment for patients with bipolar I disorder: results from a north american study of quetiapine in combination with lithium or divalproex (trial 127). *Am J Psychiatry* 2009; **166**: 476–488.
- 24 Gonzalez JM, Bowden CL, Katz MM, Thompson P, Singh V, Prihoda TJ *et al*. Development of the bipolar inventory of symptoms scale: concurrent validity, discriminant validity and retest reliability. *Int J Methods Psychiatr Res* 2008; **17**: 198–209.
- 25 Thompson PM, Gonzalez JM, Singh V, Schoolfield JD, Katz MM, Bowden CL *et al*. Principal domains of behavioral psychopathology identified by the bipolar inventory of signs and symptoms scale (BISS). *Psychiatry Res* 2010; **175**: 221–226.
- 26 Srivastava S, Ketter TA. Clinical relevance of treatments for acute bipolar disorder: balancing therapeutic and adverse effects. *Clinical Ther* 2011; **33**: B40–B48.
- 27 Singh V, Bowden CL, Gonzalez JM, Thompson P, Prihoda TJ, Katz MM *et al*. Discriminating primary clinical states in bipolar disorder with a comprehensive symptom scale. *Acta Psychiatr Scand* 2013; **127**: 145–152.



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>